

SYSTEMS AND METHODS FOR LOW BIT RATE AUDIO CODERS

This application claims priority under 35 U.S.C. 119 to US Provisional Applications No. 60/506,300 filed on September 26, 2003 which is incorporated herein
5 by reference.

Technical Field of the Invention

The present invention relates generally to audio processing and more particularly to systems and methods for use at low bit rates.

Background of the Invention

In the present state of the art, audio coders for use in coding signals representative of, for example, speech and music, for purposes of storage or transmission, perceptual models based on the characteristics of the human auditory system are typically employed
10 to reduce the number of bits required to code a given signal. In particular, by taking such characteristics into account, "transparent" coding (i.e., coding having no perceptible loss of quality) can be achieved with significantly fewer bits than would otherwise be necessary.

In such coders the signal to be coded is first partitioned into individual frames
20 with each frame comprising a small time slice of the signal, such as, for example, a time slice of approximately twenty milliseconds. Then, the signal for the given frame is transformed into the frequency domain, typically with use of a filter bank. The resulting spectral lines may then be quantized and coded.

In particular, the quantizer which is used in a perceptual audio coder to quantize
25 the spectral coefficients is advantageously controlled by a psychoacoustic model (i.e., a model based on the performance of the human auditory system) to determine masking thresholds (distortionless thresholds) for groups of neighboring spectral lines referred to as one scale factor band. The psychoacoustic model gives a set of thresholds that indicate the levels of Just Noticeable Distortion (JND), if the quantization noise introduced by the
30 coder is above this level then it is audible. As long as the Signal to (quantization) Noise Ratio (SNR) of the spectral bands are higher than the Signal to Mask Ratio (SMR) the

quantization noise cannot be perceived. The spectral lines in these scale factor bands are then non-uniformly quantized and noiselessly coded (Huffman coding) to produce a compressed bit stream. The Quantizer uses different values of step sizes for different scale factor bands depending on the distortion thresholds set by a psychoacoustic block.

5 The parameter controlling the compression ratios achieved by the encoder is externally decided by a bit rate parameter, which is the data rate of an output bit stream. Depending on the mode of operation, the data rate per frame can be variable or constant or can average around a constant bit rate. For applications involving streaming at low bit rates the preferred mode of operation is one of constant bit rate.

10 In one conventional method, quantization is carried out in two loops in order to satisfy perceptual and bit rate criteria. Prior to quantization, the incoming spectral lines are raised to a power of $3/4$ (Power law Quantizer) so as to provide a more consistent SNR over the range of quantizer values. The two loops, to satisfy the perceptual and the bit rate criteria, are run over the spectral lines. The two loops consist of an outer loop
15 (distortion measure loop) and an inner loop (bit rate loop). In the inner loop, the quantization step size is adjusted in order to fit the spectral lines within a given bit rate. The above process involves modifying the step size (referred to as the global gain, as it is common for the spectrum) until the quantized spectral lines fit into a specified number of bits. The outer loop then checks for the distortion caused in the spectral lines on a band-
20 by-band basis, and increases quantization precision for bands that have distortion above JND. The quantization precision is raised through step sizes referred to as local gains. The above iterative process repeats itself until both the bit rate and the distortion conditions are met.

25 The masking thresholds are usually computed frame-by-frame and slight variations of one masking threshold from one frame to the next may lead to very different bit assignments. As a result, at low bit rates some groups of spectral coefficients may appear and disappear. This spurious energy constitutes several auditory objects, which are different from the main energy and are thus clearly perceived. These kinds of artifacts, known as "birdies", are generally encountered at low bit rates.

30 Conventional solution to quantize with minimal distortion is to employ a low pass filter. This ensures that most of the high frequency content disappears and hence the total

number of critical bands to encode comes down. This generally leads to degradation in signal quality. However, this solution does not guarantee the disappearance and appearance of the in-band frequency content, and hence does not ensure complete elimination of the birdie artifact.

5

Summary of the Invention

The present invention enhances audio quality while operating at low bit rates without introducing birdie artifacts. In one example embodiment, a perceptual audio coder uses a modified conventional two-loop approach to maintain the audio quality at medium to high bit rates and reduces occurrence of artifacts at low bit rates during quantization. In this example embodiment, the perceptual audio coder chooses quantization steps sizes based on a user specified bit rate and a perceptual priority chart for each critical band. In addition, the critical bands are preserved so as to reduce their appearance and disappearance of the critical bands and thereby reducing the occurrence of the birdie artifacts.

In an another example embodiment, a method of quantizing an audio signal includes iteratively incrementing a quantization step size of each scale factor band of a current audio frame. The number of bits consumed in quantizing spectral lines in the scale factor bands in the current frame is then compared to a specified bit rate. Scale factor bands are then checked to determine whether they are at a vanishing point. The quantization step sizes of these scale factor bands are then frozen and quantization stops, i.e., exited from quantization, when the number of bits consumed in quantizing the spectral lines in the scale factor bands is at or below the specified bit rate.

25

Brief Description of the Drawings

FIG. 1 is a flowchart illustrating a two-loop quantization technique.

FIG. 2 is a flowchart illustrating a two-loop quantization technique using a psychoacoustic model.

FIG. 3 is a block diagram illustrating an example perceptual audio coder.

FIG. 4 is an example of a suitable computing environment for implementing embodiments of the present invention.

Detailed Description of the Invention

The present subject matter provides a modified two-loop quantization technique that maintains audio quality at medium to high bit rates while reducing artifacts at low bit rates. In one example embodiment, the technique saves vanishing bands by stealing bits from surviving bands to reduce the artifacts at low bit rates.

In the following detailed description of the embodiments of the invention, reference is made to the accompanying drawings that form a part hereof, and in which are shown by way of illustration specific embodiments in which the invention may be practiced. These embodiments are described in sufficient detail to enable those skilled in the art to practice the invention, and it is to be understood that other embodiments may be utilized and that changes may be made without departing from the scope of the present invention. The following detailed description is, therefore, not to be taken in a limiting sense, and the scope of the present invention is defined only by the appended claims.

The terms "coder" and "encoder" are used interchangeably throughout the document. Also, the terms "bands", "critical bands", and "scale factor bands" are used interchangeably throughout the document. In addition, the terms "perceptual priority chart", "perceptual relevance", and "priority chart" are used interchangeably throughout the document.

FIG. 1 is a flowchart illustrating an example embodiment of a method 100 of a modified two-loop quantization technique according to the present subject matter. At 110, the method 100 in this example embodiment forms critical bands by grouping spectral lines in a received current frame. In some embodiments, an audio signal is partitioned into successive frames. Sets of neighboring spectral lines in each frame are then grouped to form critical bands.

At 115, an initial quantization step is assigned to each formed critical band. In some embodiments, the initial quantization step size of each formed critical band is set to a value of '0'. In either case, the initial step quantization step size is set such that none of the formed critical bands are lost.

At 120, the grouped sets of neighboring spectral lines are quantized according to the initially set quantization step sizes and number of bits consumed in each critical band is determined as a result of the quantization.

At 125, the critical bands are checked to determine whether the number of bits consumed by the critical bands to quantize the spectral lines in the critical bands is at or below a user specified bit rate. In some embodiments, the user specified bit rate can be a predetermined bit rate. In these embodiments, the number of bits consumed in each critical band is checked to determine whether they are at or below the user specified bit rate.

At 130, quantization step sizes of all the critical bands are frozen and exited from the quantization of the current frame if it is determined that the number of bits consumed is at or below the user specified bit rate at 125. At 135, quantization step size of each critical band is incremented by a predetermined quantization step size if it is determined that the number of bits consumed is above the user specified bit rate at 125. In some embodiments, the predetermined quantization step size is computed as a function of previous and current frame characteristics, such as the bit rates, the quantization step sizes, and whether the quantization step sizes are incremented up or down.

At 140, the critical bands in the current frame are checked to determine whether one or more critical bands are at a vanishing point. The vanishing point refers to a quantization value of substantially close to '0' (i.e., it is a point at which any increase in the quantization step size can result in a quantized value of '0'). Beyond this point the critical band can be lost. In some embodiments, an initial or starting quantization step size is assigned to each critical band based on a perceptual priority chart. In other embodiments, the initial quantization step size of each critical band is set to a value of '0'. The method 100 goes to act 125 and repeats acts 125 -140 if it is determined that none of the critical bands are at the vanishing point at 140.

At 145, quantization step sizes of the one or more critical bands that are at the vanishing point are frozen if it is determined that the one or more critical bands are at the vanishing point at 140. At 150, the spectral lines in each of the remaining critical bands are quantized and the number of bits consumed to quantize the spectral lines in the remaining critical bands is determined.

At 155, the number of bits consumed by the spectral lines in the remaining critical bands is checked to determine whether the number of bits consumed is at or below the user specified limit. At 160, quantization step sizes of all the remaining critical bands are

frozen and exited from the quantization of the current frame if it is determined that the number of bits consumed is at or below the user specified bit rate at 155. At 165, quantization step sizes of the remaining critical bands are incremented by the predetermined quantization step size if it is determined that the number of bits consumed to quantize the spectral lines in the remaining critical bands are above the user specified bit rate at 155.

At 170, the remaining critical bands are checked to determine whether all the critical bands are at the vanishing point. At 170, the method 100 goes to act 145 and repeats acts 145 - 170 if it is determined that not all the remaining critical bands are at the vanishing point, i.e., one or more of the remaining critical bands are at the vanishing point.

At 175, the remaining critical bands are compared with a perceptual priority chart if it is determined that all the critical bands are at the vanishing point at 170. At 180, one or more of the critical bands having a low perceptual priority are dropped as a function of the comparison at 175. In these embodiments, the one or more critical bands that do not affect quality of the audio signal, based on a perceptual relevance, are dropped during quantization.

At 185, the method 100 again checks to determine whether the number of bits consumed to quantize the spectral lines in the remaining critical bands is at or below the user specified bit rate. The method 100 goes to act 180 and repeats acts 180 - 185 if it is determined that the number of bits consumed is above the user specified bit rate at 185. At 190, quantization step sizes of all the remaining critical bands are frozen and exited from the quantization of the current frame if it is determined that the number of bits consumed is at or below the user specified bit rate at 185.

FIG. 2 is a flowchart illustrating an example embodiment of a method 200 of a modified two-loop quantization technique using a psychoacoustic model according to the present subject matter. The method 200 is similar to method 100 except that the method 200 includes modified acts 215, 235, 245, 265, and 275 based on the use of the psychoacoustic model.

At 215, in the method 200 and as shown in FIG. 2, quantization step sizes for the critical bands are set based on a perceptual model and a first perceptual priority chart is

formed using the set critical bands. At 235, quantization step sizes for the critical bands are incremented based on the formed first perceptual priority chart if it is determined that the number of bits consumed by the spectral lines in the critical bands during quantization is above the user specified bit rate at 225.

5 At 245, quantization step sizes of the one or more critical bands that are at the vanishing point are frozen and a second perceptual priority chart is formed by removing the one or more critical bands, that are at the vanishing point, from the first perceptual priority chart if it is determined that the quantization step sizes of the one or more critical bands are at the vanishing point at 240. At 265, quantization step size of each remaining
10 critical band is incremented according to the formed second perceptual priority chart if it is determined that the number of bits consumed by the spectral lines in the remaining critical bands during quantization is above the user specified bit rate at 255. At 275, the remaining critical bands are compared with the first perceptual priority chart if it is determined that the quantization step sizes in all the remaining critical bands are at the
15 vanishing point at 270.

 Although the above methods 100 and 200 include acts that are arranged serially in the exemplary embodiments, other embodiments of the present subject matter may execute two or more blocks in parallel, using multiple processors or a single processor organized as two or more virtual machines or sub-processors. Moreover, still other
20 embodiments may implement the blocks as two or more specific interconnected hardware modules with related control and data signals communicated between and through the modules, or as portions of an application-specific integrated circuit. Thus, the above exemplary process flow diagrams are applicable to software, firmware, and/or hardware implementations.

25 Referring now to FIG. 3, there is illustrated an example embodiment of an audio coder 300 according to the present subject matter. The audio coder 300 includes an input module 310, a time-to-frequency transformation module 320, a psychoacoustic analysis module 330, and a bit allocator 340. The audio coder 300 further includes an encoder 350 coupled to the time-to-frequency transformation module 320 and the psycho acoustic
30 analysis module 330. As shown in FIG. 3, the encoder 350 includes an inner loop module

354 and an outer loop module 356. Further, the audio coder 300 shown in FIG.3, includes a bit stream multiplexer 370 coupled to the encoder 350 and the bit allocator 340.

In operation, in one example embodiment, the input module 310 receives an audio signal representative of, for example, speech and music, for purposes of storage or
5 transmission. Perceptual models are based on characteristics of the human auditory system typically employed to reduce the number of bits required to code a given signal. In particular, by taking such characteristics into account, "transparent" coding (i.e., coding having no perceptible loss of quality) can be achieved with significantly fewer bits than would otherwise be necessary. The input module 310 in such cases partitions the
10 received audio signal into individual frames, with each frame comprising a small time slice of the signal, such as, for example, a time slice of approximately twenty milliseconds.

The time-to-frequency transformation module 320 then receives each frame and transforms into the frequency domain, typically with the use of a filter bank, including
15 spectral lines/coefficients. Further, the time-to-frequency module 320 forms critical bands by grouping neighboring spectral lines, based on critical bands of hearing, within each frame.

The psychoacoustic module 330 then receives the audio signal from the input module 310 and determines the effects of the psychoacoustic model. The bit allocator 340
20 then estimates the bit demand based (i.e., the number of bits requested by the encoder 350 to code a given frame) based on the determined psychoacoustic model. The bit demand typically varies, having a large range, from frame to frame. The bit allocator 340 then allocates number of bits that can be given to the encoder 350 based on a predetermined bit rate to code the frame.

25 The inner loop module 354 then determines whether the number of bits consumed by the spectral lines in the critical bands in the current frame during quantization is at or below a user specified bit rate. The inner loop module 354 freezes quantization step sizes in all the critical bands when the number of bits consumed is at or below the user specified bit rate.

30 The outer loop module 356 increments quantization step sizes of the critical bands by a predetermined quantization step size when the number of bits consumed is above the

user specified bit rate. The outer loop module 356 then determines whether the quantization step sizes in one or more critical bands are at a vanishing point. The outer loop module 356 freezes the quantization step sizes in the one or more critical bands when the quantization step sizes in the one or more critical bands are at the vanishing point.

5 The outer loop module 356 quantizes spectral lines of remaining critical bands that are not at the vanishing point. The inner loop module 354 then determines whether number of bits consumed by the spectral lines in the remaining critical bands during quantization is at or below the user specified bit rate. The outer loop module 356 then freezes quantization step sizes in all the remaining critical bands and exits the quantization of the current frame
10 when the number of bits consumed is at or below the user specified bit rate.

 The outer loop module 356 increments quantization step sizes of the remaining critical bands by the predetermined quantization step size. The outer loop module 356 then determines whether the remaining critical bands are at the vanishing point.

 The outer loop module 356 then increments quantization step sizes of all the
15 critical bands and repeats the above-described functions until the user specified bit rate is met when the quantization step sizes of all the critical bands are not at the vanishing point. The outer loop module 356 compares the critical bands with a perceptual priority chart when the quantization step sizes of all the critical bands are at the vanishing point. The outer loop module 356 then drops the one or more critical bands having a lower perceptual
20 quality as a function of the comparison. The inner loop module 354 then determines whether the number of bits consumed by the spectral lines during quantization in the remaining critical bands is at or below the user specified bit rate in the current frame. The outer loop module 356 then freezes the quantization step sizes of all the remaining critical bands when the number of bits consumed by the remaining critical bands is at or below the
25 user specified bit rate. The outer loop module 356 drops one or more critical bands until the user specified bit rate is met when the number of bits consumed by the remaining critical bands are above the user specified bit rate. The operation of the encoder 350 is explained in more detail with reference to FIGS. 1 and 2.

 Various embodiments of the present invention can be implemented in software,
30 which may be run in the environment shown in FIG. 4 (to be described below) or in any other suitable computing environment. The embodiments of the present invention are

operable in a number of general-purpose or special-purpose computing environments. Some computing environments include personal computers, general-purpose computers, server computers, hand-held devices (including, but not limited to, telephones and personal digital assistants of all types), laptop devices, multi-processors, microprocessors, set-top boxes, programmable consumer electronics, network computers, minicomputers, mainframe computers, distributed computing environments and the like to execute code stored on a computer-readable medium. The embodiments of the present invention may be implemented in part or in whole as machine-executable instructions, such as program modules that are executed by a computer. Generally, program modules include routines, programs, objects, components, data structures, and the like to perform particular tasks or to implement particular abstract data types. In a distributed computing environment, program modules may be located in local or remote storage devices.

FIG. 4 shows an example of a suitable computing system environment for implementing embodiments of the present invention. FIG. 4 and the following discussion are intended to provide a brief, general description of a suitable computing environment in which certain embodiments of the inventive concepts contained herein may be implemented.

A general computing device, in the form of a computer 410, may include a processing unit 402, memory 404, removable storage 412, and non-removable storage 414. Computer 410 additionally includes a bus 405 and a network interface (NI) 401.

Computer 410 may include or have access to a computing environment that includes one or more input elements 416, one or more output elements 418, and one or more communication connections 420 such as a network interface card or a USB connection. The computer 410 may operate in a networked environment using the communication connection 420 to connect to one or more remote computers. A remote computer may include a personal computer, server, router, network PC, a peer device or other network node, and/or the like. The communication connection may include a Local Area Network (LAN), a Wide Area Network (WAN), and/or other networks.

The memory 404 may include volatile memory 406 and non-volatile memory 408. A variety of computer-readable media may be stored in and accessed from the memory elements of computer 410, such as volatile memory 406 and non-volatile memory 408,

removable storage 412 and non-removable storage 414. Computer memory elements can include any suitable memory device(s) for storing data and machine-readable instructions, such as read only memory (ROM), random access memory (RAM), erasable programmable read only memory (EPROM), electrically erasable programmable read only memory (EEPROM), hard drive, removable media drive for handling compact disks (CDs), digital video disks (DVDs), diskettes, magnetic tape cartridges, memory cards, Memory Sticks™, and the like; chemical storage; biological storage; and other types of data storage.

“Processor” or “processing unit,” as used herein, means any type of computational circuit, such as, but not limited to, a microprocessor, a microcontroller, a complex instruction set computing (CISC) microprocessor, a reduced instruction set computing (RISC) microprocessor, a very long instruction word (VLIW) microprocessor, explicitly parallel instruction computing (EPIC) microprocessor, a graphics processor, a digital signal processor, or any other type of processor or processing circuit. The term also includes embedded controllers, such as generic or programmable logic devices or arrays, application specific integrated circuits, single-chip computers, smart cards, and the like.

Embodiments of the present invention may be implemented in conjunction with program modules, including functions, procedures, data structures, application programs, etc., for performing tasks, or defining abstract data types or low-level hardware contexts.

Machine-readable instructions stored on any of the above-mentioned storage media are executable by the processing unit 402 of the computer 410. For example, a computer program 425 may comprise machine-readable instructions capable of enhancing audio quality of an audio signal when encoding at low bit rates according to the teachings and herein described embodiments of the present invention. In one embodiment, the computer program 425 may be included on a CD-ROM and loaded from the CD-ROM to a hard drive in non-volatile memory 408. The machine-readable instructions cause the computer 410 to encode an audio signal by using a modified two-loop approach that ensures maintenance of audio quality at medium to high bit rates and avoid artifacts at low bit rates according to some embodiments of the present invention.

The above description is intended to be illustrative, and not restrictive. Many other embodiments will be apparent to those skilled in the art. The scope of the invention should therefore be determined by the appended claims, along with the full scope of equivalents to which such claims are entitled.